Previously uncharacterized rectangular bacteria in the dolphin mouth

Natasha K. Dudek[1,2,9], Jesus G. Galaz-Montoya[3], Handuo Shi[3,4], Megan Mayer[5,10], Cristina Danita[3], Arianna I. Celis[4], Gong-Her Wu[3], Barry Behr[6], Kerwyn Casey Huang[3,4,7], Wah Chiu[3,4,5], David A. Relman[1,4,7,8,11,*]

[1]Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

[2]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95064, USA

[3]Department of Bioengineering, Stanford University, Stanford CA 94305, USA

[4]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA

[5]Division of CryoEM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

[6]Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, CA 94305, USA

[7]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

[8]Infectious Diseases Section, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304, USA

[9]Present address: Department of Computer Science, Mila-McGill University, Montreal, QC H3A 0G4, Canada

[10]Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

[11]Lead Contact

*Correspondence: relman@stanford.edu

SUMMARY

Much remains to be explored regarding the diversity of host-associated microbes. Here, we report the discovery of microbial structures in the mouths of bottlenose dolphins that we refer to as rectangular cell-like units (RCUs). DNA staining revealed multiple paired bands that suggested cells in the act of dividing along the longitudinal axis. Deep sequencing of samples enriched in RCUs through micromanipulation indicated that the RCUs are bacterial and distinct from *Simonsiella*, a genus with somewhat similar morphology and division patterning found in oral cavities of animals. Cryogenic transmission electron microscopy and tomography showed that RCUs are composed of parallel membrane-bound segments, likely individual cells, encapsulated by an S-layer-like periodic surface covering. RCUs displayed pilus-like appendages protruding as bundles of multiple threads that extend parallel to each other, and splay out at the tips and/or intertwine, in stark contrast to all known types of bacterial pili that consist of single, hair-like structures. These observations highlight the diversity of novel microbial forms and lifestyles that await discovery and characterization using tools complementary to genomics such as microscopy.

KEYWORDS

dolphin, microbiota, bacterial morphology, rectangular morphology, pili, S-layer, microscopy, cryo-transmission electron microscopy (cryoEM), cryo-electron tomography (cryoET), single-cell genomics

INTRODUCTION

The earliest descriptions of the microbial world centered around the morphology and motility patterns of 'animalcules' (Leeuwenhoek, 1677). In the centuries since Leeuwenhoek's revolutionary advance, a vast diversity of microbial forms have been discovered, ranging from star-shaped bacteria in the *Stella* genus (Nikitin et al., 1966; Vasilyeva, 1985) to the multicellular fruiting bodies characteristic of Myxobacteria (Dworkin, 1999; Voelz and Reichenbach, 1969). Morphology is a biologically important characteristic, often highly conserved and molded over time by selective pressures resulting from an organism's lifestyle and environmental context (Young, 2006). Indeed, cell morphology plays an important role in motility, nutrient acquisition, cell division, and interactions with other cells, including symbioses with hosts, all of which are strong determinants of survival (Young, 2007). As such, morphological and structural studies offer an appealing route by which to glean insight into microbial life forms and the mechanisms by which species function and affect their environments. Moreover, characterizing the structures and functions of the diverse range of microbes in uncharted branches of the tree of life provides an opportunity to broaden our understanding of evolution and may result in myriad applications in biotechnology and medicine, exemplified by the development of optogenetics (Fenno et al., 2011) and CRISPR-based gene editing (Ishino et al., 2018).

Genomics serves as a powerful lens through which to describe the microbial world. In recent years, metagenomic and single-cell genomic analyses have substantially increased the number of known microbial phylum-level lineages, by a factor of nearly four in the bacterial domain (Anantharaman et al., 2016; Brown et al., 2015; Castelle et al., 2015; Rinke et al., 2013). Sequencing the genomes of newly discovered organisms has

led to the discovery of new functional systems, types of protein variants, and lifestyles

(Brown et al., 2015; Burstein et al., 2017; Donia et al., 2014; Dudek et al., 2017; Wrighton

et al., 2012), illustrating the correlation between phylogenetic diversity and functional

potential (Wu et al., 2009). However, the applicability of such approaches is mostly limited

to proteins and regulatory systems homologous to those of well-characterized organisms;

the prediction of phenotypes and functions that are truly novel and/or whose genetic basis

is unknown generally requires complementary knowledge. Given the recalcitrance of the

majority of the microbial species on Earth to laboratory culturing (Hug et al., 2016),

microscopy offers an appealing route by which to study novel morphological and

functional properties of uncultured lineages. Recent advances in cryo-electron

microscopy and tomography have allowed for intact bacterial cells to be imaged in three

dimensions at nanometer resolution (Pilhofer et al., 2010), leading to important advances

in the discovery and characterization of new microbial structures (Moissl et al., 2005;

Tocheva et al., 2010).

Despite the diversity of microbial cell shapes, rectangular structures are a rarity, with a

poorly understood genetic basis. Such structures are of two types: individual cells that

are rectangular, and cell aggregates that form rectangles. To the best of our knowledge,

the discovery of non-eukaryotic rectangular cells has thus far been restricted to the family

Halobacteriaceae, which consists of halophilic Archaea. Known rectangular cells from this

family include *Haloquadratum walsbyi* (Walsby, 1980), *Haloarcula quadrata* (Oren et al.,

1999), and members of the pleomorphic genus *Natronrubrum* (Xu et al., 1999). Additional

rectangular cells believed to be bacterial or archaeal have been discovered in high salinity

environments but not taxonomically identified (Alam et al., 1984). Amongst eukaryotic

microorganisms, diatoms can have a rectangular appearance (at least when visualized in

two dimensions), although these cells are cylindrical rather than rectangular prisms

(Horner, 2002). A variety of bacteria form rectangular cell clusters, such as sheets of

coccoid bacteria (for example, *Thiopedia rosea* and the genus *Merismopedia* (Zinder and

Dworkin, 2006)), cuboidal structures of coccoid bacteria (for example, the genera *Sarcina*

and *Eucapsis* (Zinder and Dworkin, 2006)), rectangular chains of filamentous bacteria (for

example, the genus *Simonsiella* (Hedlund and Kuhn, 2006)), and rectangular trichomes

formed by disc-shaped bacteria (for example, *Oscillatoria limosa* and other cyanobacteria

(Zinder and Dworkin, 2006)).

Also rare in the microbial world are cells that diverge from the typical pattern of cell

division along a transverse axis. One spectacular example is the nematode symbiont

Robbea hypermnestra, which divides along its longitudinal axis (Leisch et al., 2016).

During division, *R. hypermnestra* cells form rectangles as they divide (as they are two

adjacent cells) and form a protective sheath around the nematode. This division

patterning is thought to preserve attachment to the host (Leisch et al., 2016). Similarly,

members of the genus *Simonsiella* divide longitudinally, which is thought to help with

adherence to human epithelial cells in the oral cavity (McCowan et al., 1979). Although

both of these bacterial taxa are uncultivated, deep insights about their biology have been

gleaned from microscopy-based studies. Such insight into the reproductive methods of

diverse bacteria is essential for building a comprehensive understanding of cell biology.

Previous studies using 16S rRNA gene amplicon sequencing and genome-resolved

metagenomics found that the mouths of bottlenose dolphins (*Tursiops truncatus*) host a

rich collection of novel microbes and functional potential (Bik et al., 2016; Dudek et al.,

2017). This diversity includes representatives from poorly characterized bacterial and archaeal phyla (Hug et al., 2016), some of which lack cultured representatives altogether and whose biology is therefore particularly poorly understood. To investigate the morphology and potential lifestyles of these novel microbial lineages, we surveyed dolphin oral microbial communities using various microscopy techniques. We discovered unusual rectangular cell-like units (RCUs) in dolphin oral samples and characterized their cellular dimensions and DNA patterning using phase-contrast and fluorescence microscopy. Regular bands of DNA suggested that the units are sheets of individual cells, and metagenomic sequencing strongly suggested that they are bacterial. Using cryogenic transmission electron microscopy (cryoTEM) and tomography (cryoET), we characterized the envelope structure of RCUs and discovered previously unobserved surface features such as heterogeneous bundles of appendages that protrude from the ends of RCUs and splay out at the tips. These findings highlight the power of high-resolution microscopy for exploring the nature of uncultivated microbes.

RESULTS

Rectangular cells in the dolphin oral cavity are Gram-negative and contain multiple bands of DNA

We collected a total of 28 oral swab samples from the mouths of nine bottlenose dolphins (*Tursiops truncatus*) during two distinct intervals in 2012 and 2018, under the purview of the U.S. Navy Marine Mammal Program in San Diego Bay, California, USA (**Methods**). In phase-contrast images, rectangular cell-like units (RCUs) were readily apparent in 20 of the 28 samples (**Figure 1A-E**), including at least one sample from each of the nine

dolphins (**Table S1**). The RCUs resembled rectangular "pancakes" (**Movie S1)** and

exhibited Gram-negative characteristics following Gram staining (**Figure 1F**). RCUs

contained multiple parallel bands of fluorescence with DAPI staining (**Figure 1B,C,D,E**).

In some RCUs, the neighboring DNA band pairs formed "H"-like shapes (**Figure 1G**, white

arrow), suggesting two rod-shaped cells in the process of division along a longitudinal

axis with the DNA bands undergoing segregation. RCUs clustered into two morphotypes

based on the length of the DAPI-stained bands. In the "long DAPI-band" morphotype

(**Figure 1B,C**), the individual DAPI-stained bands tended to be longer than their "short

DAPI-band" morphotype counterparts (**Figure 1D,E**); the nomenclature does not refer to

the number of bands per RCU). We segmented and quantified the dimensions for the two

morphotypes (**Figure 1H**, inset), and found that the "short DAPI-band" morphotype

exhibited smaller width and length compared to the "long DAPI-band" morphotype

(**Figure 1H**). The different morphotypes may represent different cell types or taxonomic

groups (for example, strains or species), cells in different stages of development, or cells

with altered shape in response to environmental conditions. We focused on cells with long

DAPI bands, as this morphotype was more prevalent in the dolphin oral samples.

RCUs are likely bacterial and not members of the genus *Simonsiella*

Given the intriguing morphology of RCUs, we next sought to determine their taxonomy

and phylogenetic affiliations. RCUs bear qualitative resemblance to members of the

genus *Simonsiella*, which consist of rod-shaped bacteria that collectively form a

rectangular unit and are oral commensals in mammals (Hedlund and Kuhn, 2006). Thus,

we first sought to evaluate the likelihood that RCUs are members of this genus. In a re-

analysis of the Sanger clone library and 454 pyrosequencing data from a previous 16S

rRNA gene amplicon survey of gingival swab samples from 38 dolphins from the same

population (Bik et al., 2016), no *Simonsiella* amplicons were detected in any of these samples. *Simonsiella* was detected in the Sanger library from the mouth and gastric fluid of a sea lion examined in the same amplicon study, confirming that *Simonsiella* DNA can be extracted successfully with the protocol used in this and the previous study (Bik et al., 2016). We next selected three new dolphin oral samples that we visually confirmed to contain RCUs and performed 16S rRNA amplicon sequencing (**Methods**), resulting in the detection of 394 amplified sequence variants (ASVs) with 991,592 amplicon reads across the three samples (**Figure 2**). Thirty-one ASVs from 21 taxa were shared among the three dolphin oral samples and are thus potential RCU ASV candidates (**Table S2**). No *Simonsiella* amplicons were detected in these three samples.

The marine origin and rectangular nature of the RCUs also gave rise to speculation that they may be marine diatoms (for example, *Skeletonema costatum*), as cylindrical marine diatoms may appear rectangular in two dimensions. Thus, we next sought evidence of eukaryotic RCU origin. We performed fluorescence *in situ* hybridization (FISH) using labeled eukaryotic (Euk-1209) and bacterial (Eub-338) probes, the latter of which is known to hybridize with both bacteria and archaea. As controls, we cultured and included *S. costatum* and the bacterium *Escherichia coli*. The Eub-338 probe hybridized to both *E. coli* and the RCUs, while the Euk-1209 eukaryotic probe hybridized to *S. costatum* cells alone (**Figure 3**), indicating that RCUs are not eukaryotic and thus not diatoms.

With no further *a priori* hypotheses as to the specific nature of the RCUs, we pursued a variety of general approaches to shed light on their identity. First, we cultured oral samples under aerobic and anaerobic conditions in three different media used to grow diverse bacteria (Methods), hoping to enrich for RCUs. Unfortunately, RCUs were not

visible upon inspection of cultures under a microscope, indicating that the growth

requirements for RCUs are distinct from those for typical bacteria isolated from

mammalian microbiotas.

Our next strategy employed single-cell genomics. Notably, this approach largely avoids

preconceived biases about possible identity, since single-cell genomics should detect

DNA from any cell from any domain of life, assuming successful cell lysis. We used three

techniques to capture rectangular cells for genomic sequencing: laser capture

microdissection, microfluidics, and cell micromanipulation. Due to their large size

compared with other bacteria, low density, and the propensity of RCUs to stick to other

cells and to abiotic surfaces, only micromanipulation led to successful RCU capture

(**Figure S1**). In addition to four collection tubes each containing ~1-3 RCUs, four negative-

control tubes of sample fluid were collected with the micropipette without any cells visible

at the resolution of our microscope. Cell-free DNA and small non-target cells were also

likely collected along with the RCUs, given the frequent close proximity of RCUs to other

cells. DNA from RCU-positive and -negative samples was amplified using multiple

displacement amplification (MDA), co-assembled, and sorted into 18 genome bins

(**Figure 4**; **Figure S2; Tables S3, S4**; **Methods**). In addition to providing several

candidate identities for the RCUs, this experiment provided further evidence against the

RCUs being marine diatoms or *Simonsiella*, as no bins were assigned to either of these

taxonomic groups. The eukaryotic bins matched the human genome or the fungal class

Malasseziomycetes, members of which are known commensals of human skin (Limon et

al., 2017), and hence likely represent contaminants. No archaeal bins were recovered.

The lack of a consistent, shared, and abundant set of genome sequences (likely due to

contamination from different non-RCU cells and/or lack of recovery of RCU DNA from

positive samples) precluded a confident taxonomic identification of the RCUs. When we

excluded bins that were present in any negative control with ≥1% relative abundance (an

admittedly arbitrary threshold) and for which an ASV from the same taxonomic group was

detected in all three dolphin oral samples that underwent amplicon sequencing, six

candidate taxa remained: 1) phylum Bacteroidetes, genus *Tenacibaculum* (76%

complete); 2) phylum Epsilonproteobacteria, genus *Arcobacter* (76% complete); 3)

Epsilonproteobacteria, family Campylobacteraceae (14% complete); 4) phylum

Gammaproteobacteria, family Moraxellaceae (58% complete); 5) Gammaproteobacteria,

unknown family (0% complete, i.e. there were assembled scaffolds in the bin even though

they did not encode marker genes used to estimate completeness; **Figure S3**); and 6)

Gammaproteobacteria, unknown family (83% complete). The relatively well-studied

genera *Tenacibaculum* and *Arcobacter* are unlikely candidates for RCU taxonomy, and

are more likely to be contaminants. If each RCU were comprised of ~8 cells (segments)

and the co-assembly were based on samples from >4 RCUs, then the larger copy number

of RCU genomes should have resulted in better quality genome assembly, suggesting

that a true RCU bin should have greater completeness, suggesting general technical

challenges faced by genomic approaches.

CryoEM reveals nanoscale surface and internal structures of RCUs

To gain high-resolution structural insight into RCUs, we imaged dolphin oral samples

containing high densities of RCUs using cryogenic transmission electron microscopy

(cryoTEM). Low-magnification cryoTEM images revealed that each RCU consists of

seemingly paired segments organized in parallel (**Figure 5A,B**); these segments were

oriented similar to the DAPI-stained bands seen in fluorescence microscopy images (**Figure 1C,E**). Groups of segments sometimes appeared to be in the act of separating from other groups, although our static data cannot definitively say whether this appearance was reflective of cell division. Segments were surrounded by a dense, membrane-like layer under a low-density layer (**Figures 6A,B (right) and 7A; Movies S2, S3**).

We hypothesized that RCUs are most likely aggregates of cells, with each DNA-containing segment, an individual cell. The following observations support this hypothesis: 1) segments appeared to be surrounded as a group by surface structures reminiscent of plasma membranes and/or cell walls (**Figures 6A,B (right) and 7A**); 2) segments are arranged in the same geometry (**Figure 5A,B**) as the DAPI-stained bands and FISH probe-hybridized bands (**Figure 1C,E,G**); 3) appendages protruded from the surface of individual segments (**Figure 5C,D**) ; 4) RCUs often consisted of groups of variable numbers of segments that appeared to be separating from one another (**Figures 1D,E and 6B**), suggesting that the rectangular structures do not reflect an individual cell.

Dark, spherical structures that were visible in the body of the RCUs in cryoTEM images may be storage and/or lipid granules. In one tomogram, two dense spheroidal objects (**Figure 5C**) were prominently visible and measured 192 nm × 200 nm × 192 nm (volume 3.1×10^7 nm^3) and 215 nm × 220 nm × 220 nm (volume 4.4×10^7 nm^3). Vesicle-like structures were also apparent. Notably, a surface covering with a periodicity of ~7-9 nm encapsulated the RCUs (**Figures 6 and 7; Movies S2, S3**).

To obtain more detailed three-dimensional reconstructions of RCU features, we

conducted cryoET experiments. Tilt-series acquisitions were limited to the RCU periphery

since the thickness of the RCU bodies occluded the electron beam at high tilt angles. The

thickness at the RCU periphery (<1 μm from the edge) ranged from ~323 nm to ~751 nm,

with an average value of ~509 nm (n=15). Appendages that resembled pili (Hospenthal

et al., 2017; Proft and Baker, 2009) protruded from RCUs; these appendages often

consisted of parallel, hair-like structures that formed bundles and splayed out at the tips,

sometimes intertwining and/or crossing over one another (**Figures 6B and 7B,C**). The

bundles of appendages were structurally heterogeneous, with variable lengths, bundle

widths, and numbers of tips. Notably, in examining the various features within the

tomograms, we did not observe any membrane-bound organelles reminiscent of a

nucleus, in line with a non-eukaryotic identity.

For both the appendages and periodic surface covering, subtomogram averaging (Galaz-

Montoya and Ludtke, 2017) did not yield consistent maps, likely due to the thickness of

the RCUs (often >500-600 nm thick), low signal-to-noise ratio of the tomograms, and

intrinsic characteristics of the features in question, such as the high and variable curvature

of the regions with a continuous periodic surface covering. These cell surface features,

which are essential for a cell's interaction with its environment, merit future investigation

as they defy the conventional wisdom that pili-like appendages are single, hair-like

structures (Hospenthal et al., 2017; Proft and Baker, 2009), and that periodic surface

coverings (such as S-layers) encapsulate a single cell (Fagan and Fairweather, 2014;

Sleytr et al., 2007) rather than a collection of cells.

DISCUSSION

Here, we used optical microscopy, cryoTEM, and cryoET to search for novel morphological diversity within the microbiota of dolphin oral samples, which was predicted based on previous findings of novel phylogenetic diversity and functional potential in the dolphin mouth via sequencing-based studies (Bik et al., 2016; Dudek et al., 2017). Interestingly, we discovered morphologically unusual RCUs. We infer that they are endemic to the dolphin mouth given that they were consistently present in this environment: they were identified in 20 out of 28 samples investigated, including in at least one sample from all nine dolphins included in this study, and were present in samples collected during intervals six years apart. Previous studies have found that the microbiota of marine mammals is distinct from that of seawater (even that of skin, which is constantly in direct contact with seawater) (Apprill et al., 2011; Bik et al., 2016), and thus it is unlikely that RCUs are simply contaminants from seawater.

The taxonomic identification of specific cell morphotypes from complex communities can be extremely difficult, to the point that it often remains unresolved (Alam et al., 1984; Luef et al., 2015; Wanger et al., 2008). Results from DAPI staining, cryo-EM imaging, FISH, and single-cell genomics indicate that the RCUs are bacterial, and 16S rRNA amplicon sequence-based surveys and single-cell genomics suggest that RCUs are distinct from species in the genus *Simonsiella*, a bacterial taxon whose members also form rectangularly-shaped clusters of rod-like cells. Obtaining a species-level identification via sequencing-based methods will be extremely challenging for numerous reasons, such as the frequent close proximity of RCUs with other small cells that were likely mixed with RCUs during micromanipulation or recalcitrance to laboratory lysis. Importantly, we are hesitant to exclude candidate identities based on not being present in all positive RCU

samples, since technical limitations could have resulted in false negatives. For example,

a thick cell wall or obstruction preventing reagents from reaching the RCU by the

micropipette needle could have interfered with lysis of the cell membrane and impeded

DNA extraction. Conversely, a positive result in a negative control may have arisen due

to non-specific read mapping, contamination of genome bins with material from true

contaminants, or cell-free DNA. Culturing-based approaches may ultimately be the most

promising route forward for identifying RCUs, although many combinations of parameters

will likely need to be tested to find satisfactory conditions for RCU growth. If RCUs are

strictly anaerobic, culturing may be further complicated since swabs cannot be collected

from the mouths of live dolphins without exposure to the atmosphere during the collection

process. Regardless of the taxonomic identity of the RCUs, the novel structural features

that have evolved in these microorganisms are intriguing and highlight the discovery

potential for further study.

The paired nature of segments in RCUs can likely be explained by longitudinal binary

fission behavior, as seen in members of the *Simonsiella* genus and *R. hypermnestra*

(Leisch et al., 2016; Steed, 1962). In *Simonsiella*, sheets are thought to help cells stay

physically anchored in the oral cavity when rapidly shedding epithelial cells slough off

(McCowan et al., 1979). We hypothesize that the same may be true for RCUs, which

inhabit a similar environment and whose morphology may have undergone convergent

evolution due to similar evolutionary pressures. Longitudinal binary fission may be an

even more general characteristic that is selected in response to the need to form a secure

attachment to a substrate. The segments at the ends of RCUs are often shorter than

those closer to the center, suggesting that there may be a mechanism by which the growth

of segments is determined by their spatial positioning within an RCU.

CryoTEM images suggested that RCUs are encapsulated by a periodic surface covering,

which may be an S-layer or a new crystalline structure. S-layers are self-assembling,

crystalline arrays of single proteins or glycoproteins that coat the exterior of some bacteria

and archaea (Fagan and Fairweather, 2014; Sleytr et al., 2007). While their exact function

varies widely across microorganisms and is often unknown, S-layers are hypothesized to

confer beneficial functions given their high metabolic cost (the S-layer comprises up to

~20% of the total protein synthesized by cells), their ubiquity across microbes, and their

multiple evolutionary origins (Fagan and Fairweather, 2014; Sleytr et al., 2007). If

segments in RCUs correspond to individual cells, production of the periodic surface

covering may represent cooperation between cells in RCUs. Cooperative synthesis of a

single, shared periodic surface covering by multiple cells could have evolved since close

kin (other cells in an RCU) have limited dispersal ability and are therefore situated in close

physical proximity. RCUs would benefit from cooperative production of a single periodic

surface covering around a population of cells rather than around each individual segment

by reducing the surface area required to cover all cells, and such an advantage could

even have contributed to selection for aggregation. An additional and not mutually

exclusive possibility is that the periodic surface covering may help to maintain the

ultrastructure of segments within an RCU, similar to archaea such as *Thermoproteus*

tenax (Wildhaber and Baumeister, 1987).

One of the most striking features of RCUs is their pilus-like appendages. At present, there

are five characterized classes of pili in Gram-negative bacteria (Chaperone-usher, Curli

fibers, F-type, type IV, and type V) and two general types of pili in Gram-positive bacteria

(short, thin rods and longer, flexible, hair-like filaments) (Hospenthal et al., 2017; Proft

and Baker, 2009); other pilus-like appendages have been documented, such as hami in

archaea (Moissl et al., 2005). To the best of our knowledge, characterized bacterial pili

all consist of single appendages that exist as independent units. By contrast, the pilus-

like appendages that protrude from RCU segments exhibit an unusual architecture

involving heterogeneous bundles of parallel filaments that often splay out at the tips.

These observations raise the question of whether the RCU appendages represent a novel

type of assembly of pilin subunits, or are a completely distinct class of appendages.

Extensive investigation of hundreds of cryoET tomograms enabled visualization of the

structure of many RCU features at high resolution. Future studies of RCUs may benefit

from imaging with phase-plate optics that dramatically increase image contrast (Danev

and Baumeister, 2017) following specimen preparation methods that thin cells into

lamellae by focused ion beam (FIB) milling coupled with scanning electron microscopy

(SEM) at cryogenic temperatures (Marko et al., 2007; Wu et al., 2020). Such approaches

could enable a comprehensive analysis of the community of RCUs as well as higher

resolution structural details of subcellular components of interest via subtomogram

averaging.

The vast majority of microorganisms on Earth lack isolated representatives (Hug et al.,

2016). Sequencing-based analyses have proved invaluable in exploring and describing

said diversity, yet cannot be used to explore all aspects of the biology of microorganisms.

Notable blind spots in our understanding of uncultured organisms include the unique

genes and corresponding structural and functional features that have evolved within these

lineages. While the use of advanced imaging techniques to visualize microbes can

provide insight into the biology of uncultured lineages, a shift toward a more multifaceted

approach drawing on many disciplines and techniques will be required to create a comprehensive view of this biological dark matter (Castelle and Banfield, 2018; Ponomarova and Patil, 2015).

ACKNOWLEDGMENTS

We thank members of the Relman lab for insightful discussion, Katharine Ng, and Brian Yu for assistance in attempts to capture RCUs using laser capture microdissection and microfluidics, respectively, and Melissa Clark for assistance with Gram staining, Michael Schmid for analysis of the periodic surface covering on RCUs, and especially Celeste Parry and colleagues at the U.S. Navy Marine Mammal Program in San Diego, CA for collecting dolphin oral samples. This work was supported by NIH grant P41GM103832 (W.C.), a James S. McDonnell Postdoctoral Fellowship (H.S.), and the Thomas C. and Joan M. Merigan Endowment at Stanford University (D.A.R.). K.C.H. and D.A.R. are Chan Zuckerberg Biohub Investigators.

AUTHOR CONTRIBUTIONS

Conceptualization: N.K.D, K.C.H., D.A.R.

Methodology and investigation: N.K.D., J.G.G.M., H.S., M.M., C.D., G.H.W, B.B., K.C.H, W.C., D.A.R.

Formal analysis: N.K.D, J.G.G.M., H.S., C.D.

Visualization: N.K.D, J.G.G.M, H.S., C.D.

Writing - original draft: The manuscript was mainly written by N.K.D. and J.G.G.M. based on N.K.D.'s original draft, with revisions by all other authors.

Writing - review and editing: all authors.

420   Supervision: K.C.H, W.C., D.A.R.

422   **DECLARATIONS OF INTEREST**

423   The authors declare no competing interests.

 FIGURE LEGENDS

426   **Figure 1. Light microscopy reveals RCUs with multiple morphotypes and distinct**

427   **DNA banding patterns.**

428   A) Phase-contrast images of RCUs (arrow indicates one example) on the surface of

429   dolphin oral epithelial cells. See also **Movie S1**. Scale bar: 10 µm.

430   B,C) Some RCUs exhibited long bands of DAPI fluorescence. Phase-contrast image

431   is shown in (B), with fluorescence overlay in cyan in (C). Scale bars: 10 µm.

432   D,E) Other RCUs exhibited shorter DAPI bands. Phase-contrast image is shown in

433   (D), with fluorescence overlay in cyan in (E). Dark spots (arrowheads) were

organized in lines perpendicular to DAPI-stained bands. DAPI-stained bands

appeared to be organized in pairs. Scale bars: 10 µm.

F) Gram-stained RCUs display Gram-negative characteristics. Inset: Gram-stained

Bacillus subtilis (*B.s.*, Gram-positive) and *Escherichia coli* (*E.c.*, Gram-negative).

Scale bar: 10 µm.

G) Neighboring DNA band pairs in an RCU form "H"-like shapes (white arrow), likely

because the DNA bands are nucleoids segregating in a cell undergoing

longitudinal division. Scale bar: 10 µm.

H) The two RCU morphotypes have distinct distributions of length and width.

Figure 2. 16S rRNA amplicon sequencing of dolphin oral samples indicates that RCUs are not affiliated with the *Simonsiella* genus. Three dolphin oral samples visually confirmed to contain RCUs were subjected to 16S rRNA gene amplicon sequencing. All detected ASVs are shown. Each row shows ASVs with their lowest

450   taxonomic assignment (level denoted by prefix such as "s__" for species). Taxonomic

451   groups are color-coded by phylum, with the polyphyletic Proteobacteria group shown at

452   the level of Class.

Figure 3. Fluorescence in situ hybridization indicates that RCUs are bacterial rather than eukaryotic. Bacterial probe Eub-338 was labeled with AlexaFluor-488 and eukaryotic probe Euk-1209 was labeled with AlexaFluor-660. Top: phase-contrast images; middle, bottom: fluorescence images from bacterial and eukaryotic probes, respectively. Arrows indicate the relevant cells in non-axenic samples. The first column is a cell of the marine diatom *Skeletonema costatum* grown in non-axenic culture. The second column is axenic *Escherichia coli* cells. The last three columns are RCUs obtained directly from dolphin oral swabs. All RCUs were labeled with the bacterial probe, while the eukaryotic probe only labeled the marine diatom. Scale bars: 10 μm.

Domain	Phylum	Lowest taxonomic ID	Bin ID	RCU 1	2	3	4		Neg 1	2	3	4		Criteria 1	2	3
Bacteria	Gammaproteobacteria	c_Gammaproteobacteria	18													
Bacteria	Epsilonproteobacteria	g_Arcobacter	4													
Bacteria	Gammaproteobacteria	c_Gammaproteobacteria	15													
Bacteria	Bacteroidetes	g_Tenacibaculum*	11													
Bacteria	Epsilonproteobacteria	f_Campylobacteraceae*	3													
Bacteria	Gammaproteobacteria	c_Gammaproteobacteria	10													
Bacteria	Gammaproteobacteria	f_Moraxellaceae*	13													
Bacteria	Bacteroidetes	f_Flavobacteriaceae	12													
Bacteria	Fusobacteria	g_Fusobacterium	6													
Bacteria	Betaproteobacteria	f_Alcaligenaceae	16													
Bacteria	Fusobacteria	g_Oceanivirga	14													
Bacteria	Gammaproteobacteria	g_Pasteurella*	17													
Bacteria	Bacteroidetes	g_Porphyromonas*	9													
Eukaryota	Fungi	c_Malasseziomycetes	8													
Eukaryota	Metazoa	s_Human	1													
Bacteria	Gracilibacteria	p_Gracilibacteria	5													+
Bacteria	Gracilibacteria	p_Gracilibacteria	2													+
Bacteria	Actinobacteria	f_Corynebacteriaceae*	7													

Figure 4. Insights into RCU identity. The 18 bins recovered from MDA and sequencing of RCU samples collected via micromanipulator, along with the Phylum from which they are inferred to derive. For lowest taxonomic identity achieved (or Class, in the case of the polyphyletic Proteobacteria), an asterisk (*) denotes lower confidence in the assignment (Methods). The RCU panel shows relative abundances of bins in each of the four samples that contained RCUs based on visualization, color-coded as follows: green: ≥5%, yellow: ≥1% and <5%, orange: >0% and <1%, red: 0%. The negative-control panel (Neg) presents the same information for each of the samples that did not appear to contain RCUs. Criteria for gauging the likelihood of a bin deriving from the RCUs are shown. (1) Was the bin ever present in negative controls (green = no, yellow = yes but never ≥1% relative abundance, orange = yes but never ≥5%, red = yes and at least once ≥5%). (2) Of ASVs shared between the three dolphin oral samples that produced the RCU panel, was any shared ASV a putative candidate for the bin? For example, bin 14 was identified as from the family Oceanivirga; was any shared ASV a member of Oceanivirga? (green = yes, red = no). (3) Are members of a given phylum known to be Gram-negative (green

479 = yes, red = no)? A "†" denotes that members of the Gracilibacteria phylum are inferred

480 not to be Gram-negative from genomic studies (although they are not necessarily Gram-

481 positive) (Meheust et al., 2019). Highest likelihood candidates are green for all three

482 criteria.

Figure 5. CryoTEM demonstrates that RCUs consist of multiple parallel, likely paired segments and are often near other cells. Band-pass filtered and denoised low magnification (A) and images or montage cryoTEM images of RCUs on an R2/2 holey carbon TEM grid (B-E). In (A), pairs of segments are highlighted with alternating shades of blue, and sharp indentations between groups of segments are denoted with blue arrows. Dense spheroidal objects inside RCUs may be storage granules or lipid droplets (black arrow). (D, E) RCUs were often in close physical proximity to other cells. In (E), an

491   apparent small indentation (blue arrow) in the RCU periodic surface covering overlaps

492   with a non-RCU cell. Small dark spots (15 nm) in (C-E) are gold fiducial particles for

493   cryoET experiments (white arrow). Scale bars: 1 µm.

Figure 6. CryoET reveals the three-dimensional architecture of RCU components.

A, B) Examples of ~3 nm-thick slices at two different depths (left and center) from two representative RCU tomograms, and corresponding manual annotations of cellular features (right). The tomograms are thick (~500-600 nm) and as such, pilus-like appendages (yellow) were visible only at a certain depth within the tomographic volume density of the RCU. See also **Movies S2 and S3.** Blue, inner membranes; purple, periodic surface covering; green, outer membrane; red, matrix. Scale bars: 100 nm.

Figure 7. RCU surface features include appendages in heterogeneous bundles that splay out at the tips and a periodic surface covering around the entire RCU.

A) A single cryoTEM image.

B,C) CryoET tomographic slices (~3-nm thick) of an RCU. Red and orange boxes in
(C) are magnified views of bundles of appendages, and yellow arrows denote thin,
single appendages. Scale bars: 100 nm.

509   D) Representative 2D cryoTEM image at the edge of an RCU showing a periodic

510   surface covering.

511   E) Line density profiles along selected regions from the image in (D) show that the

512   spacing of the repetitive features is ~7-9 nm along a direction parallel to the RCU

513   membrane. Scale bar: 100 nm.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David Relman (relman@stanford.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Raw single-cell sequence reads and 16S rRNA sequence reads are available through NCBI BioProject database PRJNA174530 with BioSample identifiers SAMN19013666, SAMN19013667, and SAMN19013668.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Oral swab samples were obtained from bottlenose dolphins (*Tursiops truncatus*) managed by the U.S. Navy Marine Mammal Program (MMP) Biosciences Division, Space and Naval Warfare Systems Center Pacific, San Diego, USA. The earliest sample containing RCUs was collected on April 1, 2012 and the latest on February 22, 2018. Samples were obtained by swabbing the roof of the mouth, the tongue, and between the tongue and mandible. Swabs were obtained using sterile foam Catch-All sample collection swabs (Epicenter, WI, Cat. #QEC091H). The swabbing protocol adhered to the guidelines described in the CRC Handbook of Marine Mammal Medicine. The MMP is

accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) International and adheres to the national standards of the United States Public Health Service Policy on the Humane Care and Use of Laboratory Animals and the Animal Welfare Act. As required by the U.S. Department of Defense, the MMP's animal care and use program is routinely reviewed by an Institutional Animal Care and Use Committee (IACUC) and by the U.S. Navy Bureau of Medicine and Surgery. The animal use and care protocol for MMP dolphins in support of this study was approved by the MMP's IACUC and the Navy's Bureau of Medicine and Surgery (IACUC #92-2010, BUMED NRD-681). Samples were collected during 2012 and 2018.

METHOD DETAILS

Microscopy sample preparation

To separate cells from swabs, swabs were immersed in 1X PBS (~50-100 µL, depending on cell density) in microcentrifuge tubes. Tubes were vortexed vigorously for ~10 s and lightly centrifuged to remove liquid from tube caps. The resulting solution was used for microscopy.

Light microscopy

Approximately 1 µL of cell solution in PBS was spotted onto an agarose pad (1% agarose in PBS) and imaged with an Eclipse Ti-E inverted microscope with a 100X (NA: 1.4) objective (Nikon, Tokyo, Japan). To determine DNA localization, cells were stained with DAPI at a final concentration of 0.5 µg/mL prior to imaging for 5 min and imaged using emission/excitation spectra of 340/488 nm.

Gram staining

Gram staining was performed using a PREVI COLOR GRAM automated machine (bioMerieux, Marcy-l'Étoile, France).

RCU cryofixation and cryoEM/ET data acquisition

A solution of cells in PBS (2.5 µL) was applied to glow-discharged 200-mesh copper, holey-carbon Quantifoil grids (Quantifoil, Großlöbichau, Germany, Cat. #Q2100CR1) with holey carbon or gold GridFinder Quantifoil grids (Quantifoil, Großlöbichau, Germany, Cat. #LFH2100AR2), followed by application of 2 µL of 15-nm gold fiducial solution to both sides of each grid. Grids were blotted for 5 s and plunge-frozen in liquid ethane cooled by liquid nitrogen to approximately -195 °C using an EM GP Plunge Freezer (Leica, Wetzlar, Germany).

Samples were loaded into one of two microscopes: a Titan Krios G3 operated at 300 kV with an energy filter (20-eV slit width), or a Titan Krios G4 operated at 300 kV without an energy filter. Both microscopes were equipped with a K2 Summit direct electron detection device (Gatan, Pleasanton, USA) used to record micrographs. Data were acquired semi-automatically in counting mode using SerialEM (Mastronarde, 2003). CryoEM/ET imaging parameters are provided in **Table S5**.

CryoEM/ET data processing

Montages were blended and binned 4-fold or greater using IMOD's "blendmont" algorithm (Kremer et al., 1996) and normalized, band-pass filtered, rotated, and cropped for display purposes using EMAN2 (Tang et al., 2007). Fifteen out of sixteen tilt series were suitable for tomographic reconstruction in IMOD. Tilt series with sampling at 7.5 Å/pixel were

down-sampled by 2-fold and those with sampling at 3.48 Å/pixel or 3.75 Å/pixel were

down-sampled by 4-fold. Images with artifacts such as excessive charging, drifting, large

ice contamination creeping in at high tilts, or excessive thickness at high tilt were excluded

from 12 of the tilt series prior to manual gold-fiducial-based alignment; up to 13 images

were removed out of the 41 images in the original raw tilt series. Tomograms were

reconstructed using standard weighted back-projection and a SIRT-like filter (mimicking

16 iterations) and were band-pass-filtered and further binned by 2-fold in most cases for

feature annotation, segmentation, movie production, and other display purposes.

Tomogram thickness was estimated by visually identifying the smallest and largest *z*-

slices with visible RCU or ice contamination densities and converting the number of slices

to nanometers. Subtomogram averaging was attempted using EMAN2 (Galaz-Montoya

et al., 2015; Galaz-Montoya et al., 2016) for globular densities suspected to be ribosomes,

matrix densities under the outer membrane, patches of the periodic surface covering, and

regions of pilus-like appendages, but no interpretable structures with resolution better

than ~50 Å were attained. The ranges of thickness and length for the pilus-like

appendages were derived by visually scanning the slices in the tomograms for the

thinnest individual filaments and thickest bundles perceptible to the naked eye, and

measuring their dimensions in binned-by-4 tomographic slices using the measuring tape

tool of EMAN2's `e2display.py`. The repeat distance of the periodic surface covering

was measured manually in a similar fashion as the pilus-like appendages from

tomographic slices, with ~10-20 measurements from each of three tomograms displaying

at least small regions where the repeat was discernible. This quantification yielded a

range between ~6 and 10 nm, suggesting that either the layer components are flexible or

that the underlying structure can yield different apparent distances between its subunits

depending on the angle at which it is sliced. Additionally, regions showing the pattern

much more clearly in higher-magnification montage two-dimensional projection images
were cropped out, rotated to lie in a horizontal plane, filtered, and masked to compute
line-density profiles parallel to the outer membrane. This strategy allowed measurement
of the distance between consecutive peaks and/or consecutive valleys, yielding values
between ~7-9 nm.

We initially carried out tomographic annotation of three features (periodic surface
covering, lipid membranes, and pilus-like appendages) for three tomograms using
EMAN2's semi-automated two-dimensional neural network-based pipeline (Chen et al.,
2017) and performed manual clean-up of false positives in UCSF Chimera (Pettersen et
al., 2004). The output annotation probability maps from EMAN2 were turned into
segmentations by applying a threshold that was visually satisfying and multiplying the
contrast-reversed tomograms by the thresholded annotation map. The segmentations
were low-pass-filtered with EMAN2 to smooth out noise. However, since the complexity
of subcellular structures was not captured by the semi-automated annotations, we applied
a similar process to generate segmentations of five features (pilus-like appendages,
periodic surface covering, outer membrane, matrix, and inner membranes) using manual
annotations performed with IMOD. Snapshots for **Figure 6** displaying RCU features in
color as well as **Movies S2 and 3** showing segmentation results were produced with
UCSF Chimera.

Fluorescence *in situ* Hybridization

Cell cultures of axenic *Escherichia coli* MG1655 and non-axenic *Skeletonema costatum*
LB 2308 (UTEX Culture Collection of Algae at the University of Texas at Austin, Austin,
TX, USA) were prepared as controls. *E. coli* was cultured in LB broth and grown at 37 °C,

and *S. costatum* was cultured in Erdschreiber's Medium at 20 °C with a ~12 h light and

~12 h dark cycle.

Probes Bact-338 and Euk-1209 were used to identify bacterial and eukaryotic cells,

respectively. Probes were ordered from Integrated DNA Technologies (Coralville, USA)

with HPLC purification as follows: Euk-1209 labeled with Alexa Fluor 660

(GGGCATCACAGACCTG/3AlexF660N) and Bact338 labeled with Alexa Fluor 488

(GCTGCCTCCCGTAGGAGT/3AlexF488N).

Cells from controls and RCUs were collected in microcentrifuge tubes. To ensure

sufficient biomass from dolphin oral swabs, cells from four swabs were condensed into a

single tube. The FISH protocol was adapted from (Skinner et al., 2013). Cells were fixed

in 1 mL of 3.7% formaldehyde solution (800 µL DEPC-treated water, 100 µL 10X PBS,

100 µL 37% formaldehyde) for 30 min with gentle shaking at 700 rpm. Cells were then

washed twice in 1 mL of 1X PBS, and permeabilized in a mixture of 300 µL DEPC-treated

water and 700 µL 200-proof ethanol with gentle shaking at 700 rpm for 2 h. Probes were

added to 50 µL of 40% hybridization solution (5 mL DEPC-water, 1 g dextran sulfate, 3.53

mL formamide, 1 mL 2X SSC, brought to a total volume of 10 mL with DEPC-treated

water) to a final concentration of 1 µM per probe set. Cells were incubated overnight in

50 µL of hybridization buffer with FISH probes at 30 °C. Cells were washed twice using a

wash solution (2 mL 20X SSC buffer, 7.06 mL formamide, 10.94 mL DEPC-treated water)

and resuspended in 2X SSC buffer.

Imaging data were processed using FIJI v. 2.0.0 (Schindelin et al., 2012).

16S rRNA amplicon sequencing and processing

Three dolphin oral samples that were confirmed to contain RCUs with long DAPI bands were selected for 16S rRNA amplicon sequencing. DNA was extracted using the QiaAMP DNA Mini Kit (Qiagen, Valencia, USA, Cat. #51304) as described in (Bik et al., 2016). The V4 region of the 16S rRNA gene was PCR-amplified in triplicate using barcoded 515F forward primers and the 806RB reverse primer. PCR was performed in triplicate using the 5 Prime Hot Master Mix (Quantabio, Beverly, USA Cat. #2200410). PCR products were purified using the UltraClean 96 PCR Cleanup Kit (Qiagen, Valencia, USA, Cat. #12596-4) and pooled in equimolar ratio following DNA quantification using the Quant-iT dsDNA assay kit (Thermo Fisher Scientific, Waltham, USA, Cat. #Q33120). Finally, DNA was run through a Zymo Clean and Concentrate Spin Column (Zymo Research Corporation, Irvine, USA, Cat. #D4029) and further purified using the QIAquick gel extraction kit (Qiagen, Hilden, Germany, Cat. #28704). Amplicons were sequenced in a single 2x250 nt Illumina HiSeq 2500 lane at the W. M. Keck Center for Comparative Functional Genomics at the University of Illinois, Urbana-Champaign.

Demultiplexing was performed using QIIME v. 1.9.1 (Caporaso et al., 2010). Amplicon sequencing variants (ASVs) were inferred using DADA2 v. 1.6.0 (Callahan et al., 2016), following guidelines in the "Big Data Workflow" (https://benjjneb.github.io/dada2/bigdata_paired.html). Forward and reverse reads were trimmed to 245 nt and 200 nt, respectively. ASVs were inferred separately for forward and reverse reads using lane-specific error-rate profiles, and paired reads were merged. The "removeBimeraDenovo" function in DADA2 was used to identify and remove chimeras from sample datasets and taxonomic assignments were created using the DADA2 "assignTaxonomy" and "assignSpecies" functions, using RDP training set 16 as

a reference database (Cole et al., 2014). Additional stringent filtering of amplicons was

performed, using VSEARCH v. 2.8.0 to remove chimeras (Rognes et al., 2016), BLAST

v. 2.7.1 (Altschul et al., 1990) and the 16S rRNA gene set in (Schulz et al., 2017) to retain

only bacterial 16S rRNA genes (e.g. to remove chloroplast 16S rRNA genes), scripts to

remove abnormally short ASVs (<1200 bp) or those that contained N's or X's, and

Decontam v. 0.99.3 to remove contaminant ASVs (Davis et al., 2018). This pipeline

yielded a total of 394 taxa and 991,592 reads across the three samples.

Single-cell genomics

To obtain candidate identities for RCUs, we employed a single-cell genomics approach.

To limit contamination by foreign DNA, reagents, tubes, and PBS were treated with 11.4

J/cm^2 of ultraviolet light following the guidelines in (Woyke et al., 2011). RCUs were

visualized with an Olympus IX70 inverted microscope (Olympus, Waltham, USA) with a

40X objective and Hoffman modulation optics. An Eppendorf TransferMan

micromanipulator (Eppendorf, Hamburg, Germany, Cat. #5193000020) with a SAS-10

microinjector was used to capture RCUs with Polar Body Biopsy Micropipettes (30°

angled, beveled, and polished with an inner diameter of 13-15 μm) (Cooper Surgical,

Målov, Denmark, Cat. #MPB-BP-30). After an RCU or chain of RCUs was acquired, the

micropipette tip was transferred to a collection tube containing 1X PBS and crushed into

the tube to ensure the RCU(s) was deposited in the tube; this precaution was adopted

because RCUs frequently stuck to the glass micropipettes and could not be dislodged.

No dolphin cells were captured, although cell-free DNA and small, non-target cells from

the sample were likely acquired as contaminants along with RCUs based on the

propensity for attachment of other species (**Figure 5E**). Four tubes of RCUs were

collected (sample names RCU1, RCU2, RCU3, RCU4), along with four negative-control

tubes (NEG1, NEG2, NEG3, NEG4). Negative controls consisted of draws of PBS from

the same sample that did not contain any visible cells, and were otherwise treated

identically to RCU-containing samples.

DNA from each tube was amplified via multiple displacement amplification (MDA) using

the Repli-g single-cell kit (Qiagen, Hilden, Germany, Cat. #150343) according to the

manufacturer's protocol. DNA was purified using a Zymo Clean and Concentrate Spin

Column (Zymo Research Corporation, Irvine, USA, Cat. #D4013) and libraries were

prepared using the Kapa Hyper Prep Kit (Kapa Biosystems, Wilmington, USA, Cat.

#KK8504) at the W.M. Keck Center for Comparative Functional Genomics at the

University of Illinois, Urbana-Champaign. The eight libraries were sequenced using the

Illumina MiSeq 2x250 nt P2 V2 platform. RCU samples RCU1, RCU2, NEG1, and NEG2

were pooled and sequenced across a single lane that produced 11,371,243 read pairs,

and samples RCU3, RCU4, NEG3, and NEG4 were pooled and sequenced across 1.5

lanes, collectively producing 19,615,690 read pairs. Sequencing adaptors were

computationally removed at the Keck Center.

Reads from all eight samples were co-assembled using SPAdes v. 3.11.1 (Bankevich et

al., 2012) with the single cell (–sc) and careful (–careful) modes specified. A total of

61,973,866 read pairs were used for assembly, resulting in 1,406 scaffolds ≥5 kbp long

with a total length of 17,438,233 bp and an N50 of 14,592 for scaffolds ≥5 kbp long.

Protein coding genes were identified using Prodigal v. 2.6.2 (Hyatt et al., 2010). Per

scaffold average coverage was calculated by mapping reads per sample against the co-

assembly using bowtie2 v. 2.2.4 (Langmead and Salzberg, 2012), using the samtools v.

1.6 depth function (Li et al., 2009) to calculate per-base read coverage, and a custom

script to calculate average per-base read coverage per scaffold.

To determine the taxonomic identity of sequenced cells, we employed a genome-resolved

approach. Assignment of scaffolds to genome bins was performed using the

tetranucleotide frequencies of all scaffolds ≥5 kbp long over windows of 5 kbp, as

described in (Dick et al., 2009). Results were computed and visualized using the

Databionics ESOM Tools software v. 1.1 (Ultsch and Mörchen, 2005), leading to the

reconstruction of 18 genome bins (**Figure S3**). To refine bins, we removed scaffolds for

which <50% of keys were assigned to the bin. Scaffolds <5 kbp long were not binned.

The completeness and contamination per bin were assessed using CheckM v. 1.0.7

(Parks et al., 2015). To evaluate how representative binning was of the genomes that

were sequenced, we estimated the number of prokaryotic genomes expected to be

recovered by searching the metagenome assembly for a set of 16 bacterial single copy

genes (bSCGs) assumed to be present in every genome in a single copy (Hug et al.,

2013). The median number of each bSCG was 10, suggesting ~10 prokaryotic genomes

were represented in our sequencing dataset.

Taxonomic identification of bins posed a challenge since 16S/18S rRNA genes were not

reliably amplified/sequenced/assembled, and genomes were partial with few

phylogenetically informative bacterial single copy genes (Hug et al., 2013) present in the

dataset. Hence, we used BLAST v. 2.2.30 (Altschul et al., 1990) to query all protein coding

genes from each genome against the NCBI non-redundant protein database using an e-

value of 10^{-10} and taxonomic assignments were made based on the closest protein match.

Genome bin taxonomic assignments were considered highly likely if ≥50% of the top blast

hits originated from a single taxon and were considered to be plausible if <50% but ≥33%

of the top blast hits originated from a single taxon.

There are numerous approaches by which one could assess whether a bin is "present" in

a sample, each with largely arbitrary thresholds. We focused on the relative abundance

of each bin per sample (**Table S4**) as it accounts for the length of each bin and allows for

comparisons between samples with different numbers of read pairs (Dudek et al., 2017).

Attempt at culturing RCUs

Two dolphin oral samples confirmed to contain RCUs were selected for culturing efforts.

For each sample, one milliliter of sterile PBS was added to a 1.5-mL Eppendorf tube

containing the oral swab sample. Six hundred microliters of each sample were used to

inoculate 3 mL of BSTYS (Kuhn et al., 1978) (2.75% w/v Tryptic Soy Broth, 0.4% w/v

yeast extract, 1.5% w/v agar, 10% bovine serum), SHI (Tian et al., 2010), or mSHI media

(SHI supplemented with 0.9 g/L NaCl, 2.5 g/L K_2PO_4, 0.84 g/L $NaHCO_3$, 0.17 g/L $CaCl_2$,

0.04 g/L $MgCl_2$ * $6H_2O$, and 5 g/L of dextrose), in liquid form and on thin layer agar plates.

BSTYS was selected for its use in successfully culturing bacteria (*Simonsiella* specifically)

from the oral cavities of various mammals (Kuhn et al., 1978). SHI was selected because

this medium was designed to sustain high diversity communities derived from the human

oral microflora. mSHI (modified-SHI) was included as a higher-salinity version of SHI in

an attempt to further mimic the conditions that might be found in the oral cavity of dolphins.

Inoculation was repeated under anaerobic conditions in an anaerobic chamber (COY Lab

Products, Grass Lake, USA); note that all samples were unavoidably exposed to

atmospheric oxygen prior to culturing. Cultures were incubated at 37 °C to mimic the body

temperature of dolphins. No RCUs were detected on solid media or in liquid media by

visual screening under a microscope after ~24, ~48, ~72, and ~96 h.

QUANTIFICATION AND STATISTICAL ANALYSIS

Detailed descriptions of the quantitative methods used in this paper can be found in the

Results and Method Details sections. Briefly, these descriptions include the methods

used for DNA extraction, sequencing, read quality filtering and analysis, and microscopy

data acquisition.

REFERENCES

Alam, M., Claviez, M., Oesterhelt, D., and Kessel, M. (1984). Flagella and motility behaviour of square bacteria. EMBO J *3*, 2899-2903.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U.*, et al.* (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun *7*, 13219.

Apprill, A., Mooney, T.A., Lyman, E., Stimpert, A.K., and Rappe, M.S. (2011). Humpback whales harbour a combination of specific and variable skin bacteria. Environ Microbiol Rep *3*, 223-232.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D.*, et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol *19*, 455-477.

Bik, E.M., Costello, E.K., Switzer, A.D., Callahan, B.J., Holmes, S.P., Wells, R.S., Carlin, K.P., Jensen, E.D., Venn-Watson, S., and Relman, D.A. (2016). Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. Nat Commun *7*, 10516.

Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., and Banfield, J.F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. Nature *523*, 208-211.

Burstein, D., Harrington, L.B., Strutt, S.C., Probst, A.J., Anantharaman, K., Thomas, B.C.,

Doudna, J.A., and Banfield, J.F. (2017). New CRISPR-Cas systems from uncultivated

microbes. Nature *542*, 237-241.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P.

(2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat

Methods *13*, 581-583.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello,

E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I.*, et al.* (2010). QIIME allows

analysis of high-throughput community sequencing data. Nat Methods *7*, 335-336.

Castelle, C.J., and Banfield, J.F. (2018). Major New Microbial Groups Expand Diversity

and Alter our Understanding of the Tree of Life. Cell *172*, 1181-1197.

Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J.,

Frischkorn, K.R., Tringe, S.G., Singh, A., Markillie, L.M.*, et al.* (2015). Genomic expansion

of domain archaea highlights roles for organisms from new phyla in anaerobic carbon

cycling. Curr Biol *25*, 690-701.

Chen, M., Dai, W., Sun, S.Y., Jonasch, D., He, C.Y., Schmid, M.F., Chiu, W., and Ludtke,

S.J. (2017). Convolutional neural networks for automated annotation of cellular cryo-

electron tomograms. Nat Methods *14*, 983-985.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-

Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal Database Project: data and

tools for high throughput rRNA analysis. Nucleic Acids Res *42*, D633-642.

Danev, R., and Baumeister, W. (2017). Expanding the boundaries of cryo-EM with phase

plates. Curr Opin Struct Biol *46*, 87-94.

Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome *6*, 226.

Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., and Banfield, J.F. (2009). Community-wide analysis of microbial genome sequence signatures. Genome Biol *10*, R85.

Donia, M.S., Cimermancic, P., Schulze, C.J., Wieland Brown, L.C., Martin, J., Mitreva, M., Clardy, J., Linington, R.G., and Fischbach, M.A. (2014). A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell *158*, 1402-1414.

Dudek, N.K., Sun, C.L., Burstein, D., Kantor, R.S., Aliaga Goltsman, D.S., Bik, E.M., Thomas, B.C., Banfield, J.F., and Relman, D.A. (2017). Novel Microbial Diversity and Functional Potential in the Marine Mammal Oral Microbiome. Curr Biol *27*, 3752-3762 e3756.

Dworkin, M. (1999). Introduction to the myxobacteria. Prokaryotic Development, 219-242.

Fagan, R.P., and Fairweather, N.F. (2014). Biogenesis and functions of bacterial S-layers. Nat Rev Microbiol *12*, 211-222.

Fenno, L., Yizhar, O., and Deisseroth, K. (2011). The development and application of optogenetics. Annu Rev Neurosci *34*, 389-412.

Galaz-Montoya, J.G., Flanagan, J., Schmid, M.F., and Ludtke, S.J. (2015). Single particle tomography in EMAN2. J Struct Biol *190*, 279-290.

Galaz-Montoya, J.G., Hecksel, C.W., Baldwin, P.R., Wang, E., Weaver, S.C., Schmid, M.F., Ludtke, S.J., and Chiu, W. (2016). Alignment algorithms and per-particle CTF correction for single particle cryo-electron tomography. J Struct Biol *194*, 383-394.

Galaz-Montoya, J.G., and Ludtke, S.J. (2017). The advent of structural biology in situ by

single particle cryo-electron tomography. Biophys Rep *3*, 17-35.

Hedlund, B.P., and Kuhn, D.A. (2006). The genera Simonsiella and Alysiella.

Horner, R.A. (2002). A taxonomic guide to some common marine phytoplankton.

Hospenthal, M.K., Costa, T.R.D., and Waksman, G. (2017). A comprehensive guide to

pilus biogenesis in Gram-negative bacteria. Nat Rev Microbiol *15*, 365-379.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J.,

Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K.*, et al.* (2016). A new view of the tree

of life. Nat Microbiol *1*, 16048.

Hug, L.A., Castelle, C.J., Wrighton, K.C., Thomas, B.C., Sharon, I., Frischkorn, K.R.,

Williams, K.H., Tringe, S.G., and Banfield, J.F. (2013). Community genomic analyses

constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate

roles in sediment carbon cycling. Microbiome *1*, 22.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010).

Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC

Bioinformatics *11*, 119.

Ishino, Y., Krupovic, M., and Forterre, P. (2018). History of CRISPR-Cas from Encounter

with a Mysterious Repeated Sequence to Genome Editing Technology. J Bacteriol *200*.

Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer visualization of

three-dimensional image data using IMOD. J Struct Biol *116*, 71-76.

Kuhn, D.A., Gregory, D.A., Buchanan, G.E., Jr., Nyby, M.D., and Daly, K.R. (1978).

Isolation, characterization, and numerical taxonomy of Simonsiella strains from the oral

cavities of cats, dogs, sheep, and humans. Arch Microbiol *118*, 235-241.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat

Methods *9*, 357-359.

Leeuwenhoek, A.V. (1677). Observations, communicated to the publisher by Mr. Antony

van Leewenhoeck, in a dutch letter of the 9th Octob. 1676. here English'd: concerning

little animals by him observed in rain-well-sea-and snow water; as also in water wherein

pepper had lain infused. Philosophical Transactions of the Royal Society of London *12*,

821-831.

Leisch, N., Pende, N., Weber, P.M., Gruber-Vodicka, H.R., Verheul, J., Vischer, N.O.,

Abby, S.S., Geier, B., den Blaauwen, T., and Bulgheresi, S. (2016). Asynchronous

division by non-ring FtsZ in the gammaproteobacterial symbiont of Robbea hypermnestra.

Nat Microbiol *2*, 16182.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis,

G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence

Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Limon, J.J., Skalski, J.H., and Underhill, D.M. (2017). Commensal Fungi in Health and

Disease. Cell Host Microbe *22*, 156-165.

Luef, B., Frischkorn, K.R., Wrighton, K.C., Holman, H.Y., Birarda, G., Thomas, B.C.,

Singh, A., Williams, K.H., Siegerist, C.E., Tringe, S.G.*, et al.* (2015). Diverse uncultivated

ultra-small bacterial cells in groundwater. Nat Commun *6*, 6372.

Marko, M., Hsieh, C., Schalek, R., Frank, J., and Mannella, C. (2007). Focused-ion-beam

thinning of frozen-hydrated biological specimens for cryo-electron microscopy. Nat

Methods *4*, 215-217.

Mastronarde, D.N. (2003). SerialEM: a program for automated tilt series acquisition on

Tecnai microscopes using prediction of specimen position. Microscopy and Microanalysis

9, 1182-1183.

McCowan, R.P., Cheng, K.J., and Costerton, J.W. (1979). Colonization of a portion of the

bovine tongue by unusual filamentous bacteria. Appl Environ Microbiol *37*, 1224-1229.

Meheust, R., Burstein, D., Castelle, C.J., and Banfield, J.F. (2019). The distinction of CPR

bacteria from other bacteria based on protein family content. Nat Commun *10*, 4173.

Moissl, C., Rachel, R., Briegel, A., Engelhardt, H., and Huber, R. (2005). The unique

structure of archaeal 'hami', highly complex cell appendages with nano-grappling hooks.

Mol Microbiol *56*, 361-370.

Nikitin, D., Vasilyeva, L., and Lokhmacheva, R. (1966). New and rare forms of soil

microorganisms. M: Nauka.

Oren, A., Ventosa, A., Gutierrez, M.C., and Kamekura, M. (1999). Haloarcula quadrata

sp. nov., a square, motile archaeon isolated from a brine pool in Sinai (Egypt). Int J Syst

Bacteriol *49 Pt 3*, 1149-1155.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015).

CheckM: assessing the quality of microbial genomes recovered from isolates, single cells,

and metagenomes. Genome Res *25*, 1043-1055.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C.,

and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research

and analysis. J Comput Chem *25*, 1605-1612.

Pilhofer, M., Ladinsky, M.S., McDowall, A.W., and Jensen, G.J. (2010). Bacterial TEM:

new insights from cryo-microscopy. Methods Cell Biol *96*, 21-45.

Ponomarova, O., and Patil, K.R. (2015). Metabolic interactions in microbial communities:

untangling the Gordian knot. Curr Opin Microbiol *27*, 37-44.

Proft, T., and Baker, E.N. (2009). Pili in Gram-negative and Gram-positive bacteria -

structure, assembly and their role in disease. Cell Mol Life Sci *66*, 613-635.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F.,

Darling, A., Malfatti, S., Swan, B.K., Gies, E.A.*, et al.* (2013). Insights into the phylogeny

and coding potential of microbial dark matter. Nature *499*, 431-437.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahe, F. (2016). VSEARCH: a

versatile open source tool for metagenomics. PeerJ *4*, e2584.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T.,

Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.*, et al.* (2012). Fiji: an open-source

platform for biological-image analysis. Nat Methods *9*, 676-682.

Schulz, F., Eloe-Fadrosh, E.A., Bowers, R.M., Jarett, J., Nielsen, T., Ivanova, N.N.,

Kyrpides, N.C., and Woyke, T. (2017). Towards a balanced view of the bacterial tree of

life. Microbiome *5*, 140.

Skinner, S.O., Sepulveda, L.A., Xu, H., and Golding, I. (2013). Measuring mRNA copy

number in individual Escherichia coli cells using single-molecule fluorescent in situ

hybridization. Nat Protoc *8*, 1100-1113.

Sleytr, U.B., Huber, C., Ilk, N., Pum, D., Schuster, B., and Egelseer, E.M. (2007). S-layers

as a tool kit for nanobiotechnological applications. FEMS Microbiol Lett *267*, 131-144.

Steed, P.A. (1962). Simonsiellaceae fam.nov.with characterization of Simonsiella crassa

and Alysiella filiformis. J Gen Microbiol *29*, 615-624.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J.

(2007). EMAN2: an extensible image processing suite for electron microscopy. J Struct

Biol *157*, 38-46.

Tian, Y., He, X., Torralba, M., Yooseph, S., Nelson, K.E., Lux, R., McLean, J.S., Yu, G.,

and Shi, W. (2010). Using DGGE profiling to develop a novel culture medium suitable for

oral microbial communities. Mol Oral Microbiol *25*, 357-367.

Tocheva, E.I., Li, Z., and Jensen, G.J. (2010). Electron cryotomography. Cold Spring

Harb Perspect Biol *2*, a003442.

Ultsch, A., and Mörchen, F. (2005). ESOM-Maps: tools for clustering, visualization, and

classification with Emergent SOM, Vol 46 (Univ.).

Vasilyeva, L. (1985). Stella, a New Genus of Soil Prosthecobacteria, with Proposals for

Stella humosa sp. nov. and Stella vacuolata sp. nov. International Journal of Systematic

and Evolutionary Microbiology *35*, 518-521.

Voelz, H., and Reichenbach, H. (1969). Fine structure of fruiting bodies of Stigmatella

aurantiaca (Myxobacterales). J Bacteriol *99*, 856-866.

Walsby, A.E. (1980). A square bacterium. Nature *283*, 69-71.

Wanger, G., Onstott, T.C., and Southam, G. (2008). Stars of the terrestrial deep

subsurface: a novel 'star-shaped' bacterial morphotype from a South African platinum

mine. Geobiology *6*, 325-330.

Wildhaber, I., and Baumeister, W. (1987). The cell envelope of Thermoproteus tenax:

three-dimensional structure of the surface layer and its role in shape maintenance. EMBO

J *6*, 1475-1480.

Woyke, T., Sczyrba, A., Lee, J., Rinke, C., Tighe, D., Clingenpeel, S., Malmstrom, R.,

Stepanauskas, R., and Cheng, J.F. (2011). Decontamination of MDA reagents for single

cell whole genome amplification. PLoS One *6*, e26161.

Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C.,

Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H.*, et al.* (2012). Fermentation,

hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science *337*,

1661-1665.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., Kunin, V.,

Goodwin, L., Wu, M., Tindall, B.J.*, et al.* (2009). A phylogeny-driven genomic

encyclopaedia of Bacteria and Archaea. Nature *462*, 1056-1060.

Wu, G.H., Mitchell, P.G., Galaz-Montoya, J.G., Hecksel, C.W., Sontag, E.M.,

Gangadharan, V., Marshman, J., Mankus, D., Bisher, M.E., Lytton-Jean, A.K.R.*, et al.*

(2020). Multi-scale 3D Cryo-Correlative Microscopy for Vitrified Cells. Structure *28*, 1231-

1237 e1233.

Xu, Y., Zhou, P., and Tian, X. (1999). Characterization of two novel haloalkaliphilic

archaea Natronorubrum bangense gen. nov., sp. nov. and Natronorubrum tibetense gen.

nov., sp. nov. Int J Syst Bacteriol *49 Pt 1*, 261-266.

Young, K.D. (2006). The selective value of bacterial shape. Microbiol Mol Biol Rev *70*,

660-703.

Young, K.D. (2007). Bacterial morphology: why have different shapes? Curr Opin

Microbiol *10*, 596-600.

Zinder, S.H., and Dworkin, M. (2006). Morphological and physiological diversity. The

prokaryotes A handbook on the biology of bacteria, 3rd edn Springer, New York, 185-

220.

1001